Title
The Sphagnum Genome Project

ISBN
9780128011027

Authors
Shaw, AJ
Schmutz, J
Devos, N
et al.

Publication Date
2016

DOI
10.1016/bs.abr.2016.01.003

Peer reviewed

Genomes of photosynthetic organisms and fungi
(Elsevier book series on Advances in Botanical Research, edited by Jean-Pierre Jacquot and Pierre Gadal)
Volume "Genomes and Evolution of Charophytes, Bryophytes, Club Mosses and Ferns"
edited by Stefan A. Rensing

Running title: *Sphagnum* as a model for ecological and evolutionary genomics

The *Sphagnum* genome project: a new model for ecological and evolutionary genomics

A. Jonathan Shaw[1], Jeremy Schmutz[2,3], Nicolas Devos[1], Shengqiang Shu[3], Alyssa A. Carrell[1,4], David J. Weston[4]

Affiliation:

[1] Department of Biology, Duke University, Durham, NC 27708

[2] HudsonAlpha Institute of Biotechnology, Huntsville, AL 35806

[3] Department of Energy Joint Genome Institute, Walnut Creek, CA 94598

[4] Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Abstract

The inception of the *Sphagnum* genome project marks the first plant-based sequencing project aimed specifically at carbon cycling genomics in a plant system relevant to ecological and evolutionary genomics. *Sphagnum* provides considerable intra- and inter - specific variation at the nucleotide level, and in physiology, morphology, net production, decomposition and carbon accumulation (peat formation). Together with their large number of species, their diversity in mating patterns, clear patterns of niche differentiation, *Sphagnum* provides an exceptionally valuable complement to *Physcomitrella patens* and *Ceratodon purpureus* as moss models for genomic research. Here we review the organismal biology of *Sphagnum* including phylogeny, life cycle, mating systems, ecology and niche differentiation. We include the current state of *Sphagnum* genomic resources, in vitro methods and germplasm. A use-case is provided to address questions concerning epigenetics and reproduction.

Key words: carbon cycling, ecological genomics***,*** peatmoss, *Sphagnum*

Introduction – *Sphagnum*, or more broadly, phylum Bryophyta (mosses), class Sphagnopsida (peatmosses), represents a valuable complement to *Physcomitrella patens* and *Ceratodon purpureus* as moss models for genomic research. The special value of peatmosses derives from (1) their early diverging phylogenetic position within the mosses, (2) the speciose nature of the peatmoss clade, (3) well-documented patterns of niche differentiation among *Sphagnum* species, and (4) the ecological importance of peatmosses for global carbon cycle dynamics. The *Sphagnum* genome project was formulated to take advantage of these unique features, which are briefly discussed below.

The Sphagnopsida comprise an early diverging lineage within the early diverging land moss lineage — Sometime in the early Paleozoic Era, perhaps during the Ordovician Period (Rubenstein et al. 2010, Wellman et al. 2003, 2010), green algal ancestors of the embryophyte land plant clade colonized terrestrial environments and diversified rapidly (Kendrick & Crane 1997, Steemens et al. 2009,). Life on land carried with it new abiotic challenges, and also new opportunities, including novel relationships with other terrestrial organisms such as fungi, eukaryotic protists, and prokaryotic microbes (Redecker et al. 2000, Knack et al. 2015). Terrestrial environments presented unprecedented stresses, especially in terms of water economy, nutrient acquisition, maintaining structure in high-gravity ecosystems, and elevated levels of UV radiation (Graham et al. 2003, Waters 2003). Responses to these novel environmental pressures led to monumental changes in plant phenotypes, including morphological, developmental, and

biochemical/metabolic features, and to the underlying genomic bases for traits necessary to survive on land. The nonvascular embryophyte clades (i.e., those plants lacking phloem and xylem), comprising the mosses (Bryophyta), liverworts (Marchantiophyta), and hornworts (Anthocerophyta), have long been considered the earliest groups of embryophyte land plants (Haeckel 1868, Bower 1890) based on their haploid-dominant life cycles and morphological features, and more recently this hypothesis has been strongly supported by numerous molecular phylogenetic studies (Qui 2006, Wickett et al. 2014, among others).

Comparative genomics has elucidated that many components of the molecular “toolkit” that facilitated the diversification of land plants evolved early in embryophyte evolution and can be detected in bryophytes (e.g., Floyd & Bowman 2007, Viaene et al. 2014). These observations indicate that homologues of many of the genes critical for plant development were present in the earliest embryophyte land plants; i.e., the bryophytes. Although these genes, or similar ancestral versions of modern genes, evolved early in the history of land plants, in many cases their functions have likely changed. Many good examples exist of genes present in early land plants being co-opted for new functions in more recently derived lineages (Rensing et al. 2008). In contrast to *Physcomitrella patens* and *Ceratodon purpureus*, which belong to the more derived group of peristomate mosses, *Sphagnum* evolved earlier in moss evolution (Chiang & Graham 2011) and as such, provides inferences about early genomic traits in the Bryophyta. This is true notwithstanding the fact that some phenotypic traits and genomic features characterizing *Sphagnum* evolved within the Sphagnopsida lineage

subsequent to their divergence from other mosses.

The Sphagnopsida: a speciose clade of mosses — The Sphagnopsida comprise a class of the phylum Bryophyta, and include four genera in three families (estimated numbers of species in each genus in parentheses): Sphagnaceae: *Sphagnum* (250-450); Flatbergiaceae: *Flatbergium* (2), Ambuchananiaceae: *Ambuchanania* (1), *Eosphagnum* (1) (Shaw et al. 2010a). *Ambuchanania, Eosphagnum*, and *Flatbergium* occur in the Southern Hemisphere (although *Flatbergium sericeum* reaches southern China) and none of them occur in peatlands but rather on moist to wet rocks and soils. *Sphagnum*, by far the largest genus of Sphagnopsida, dominates mire habitats (bogs and fens) in the Northern Hemisphere but also includes tropical and Southern Hemisphere species. It is difficult to accurately estimate the number of species in *Sphagnum* because most taxa outside the boreal zone are poorly known and even within the boreal peatmoss flora, species delimitation is a work in progress. Recent work has demonstrated that some common species include multiple, phylogenetically distinct lineages that may warrant individual species status (e.g., the common peat-forming species, *S. fuscum*, includes two color morphs (Andrus, 2006) that are highly distinct genetically/phylogenetically [Kyrkjeeide et al. in press]). Allopolyploid speciation has contributed to confusing systematic patterns in *Sphagnum*; most species are gametophytically haploid but both diploid and triploid taxa are also known (Karlin et al. 2009; Ricca & Shaw 2010). Systematic problems are directly relevant to work on peatmoss genomics and ecology. For example, *S. magellanicum* is generally considered to be one of

the most prominent peatmosses of wetlands throughout the Northern Hemisphere, including Alaska. Recent work (Kyrkjeeide unpublished) however, has shown that *S. magellanicum* is in fact known from only a couple sites in Alaska, and the dominant peat-forming species that has been included in numerous ecological studies as *S. magellanicum* is in fact the related allopolyploid, *S. alaskense*, which has *S. magellanicum* as one of its parents. Complexes of closely related species, most of which include one or more allopolyploids, occur in all four of the large subgenera of *Sphagnum* and clarifying these systematic problems are critical for any future genomic research on this group.

The Sphagnopsida differ from all other mosses in many developmental and morphological features. Their unique features include especially: (1) spores give rise to thallose (rather than filamentous) protonemata (shared with a few other mosses), (2) mature gametophytes typically lack rhizoids or other structures for substrate attachment, (3) gametophores have fasciculate branching with spreading and pendent branches in each fascicle, (4) dimorphic leaf cells with large, hyaline cells, dead and empty at maturity, enclosed in a network of smaller, narrow, photosynthetic chlorophyllose cells, (5) sporophytic seta lacking; capsules (sporangia) borne on pseudopodia of gametophytic origin, (6), sporogenous tissue within sporangia derived from embryonic amphithecial layers rather than endothecial tissue as in other mosses.

Ecology and niche differentiation within Sphagnum — Species of *Sphagnum* not only dominate northern peatland habitats in terms of biomass

and species richness, they play an active roll in the formation of their own microenvironments (van Breeman 1995). Boreal peatlands have long served as a model for research in community ecology because they are relatively simple systems in terms of numbers of species and community structure (compared to many tropical biomes, for example). Nevertheless, as many as 20 or more *Sphagnum* species commonly co-occur within peatland communities of Scandinavia and northern North America. There may be no other genus of plants in which so many species occur sympatrically within a single community.

A huge body of literature exists documenting niche differentiation among northern *Sphagnum* species relative to abiotic microhabitat gradients (see Rydin & Juglum 2006). Species sort along several gradients including cation concentrations and pH, and especially height above the water table (Fig. 1). Most well-developed peatlands exhibit a pronounced “hummock-hollow” structure and *Sphagnum* species occupy characteristic positions relative to these micro-topographic features. Interspecific variation along the hummock-hollow gradient (i.e., height above the water table) is related to phylogeny (Johnson et al. 2015) and the rate of evolution relative to this gradient is heterogeneous; interspecific differentiation among hummock species has generally evolved more quickly than among hollow species. More complex evolutionary patterns occur relative to ionic and pH gradients. The evolution of niches relative to pH may generally be constrained in *Sphagnum* by stabilizing selection, but exceptional rate changes appear to characterize some species in the subg. *Sphagnum* (Johnson & Shaw 2015). Other examples of rapid niche changes, for example, relative to substrate

electrical conductivity, have occurred at or near the tips of the *Sphagnum* phylogenetic tree; i.e., relatively recently in peatmoss evolution.

Mating systems also vary across *Sphagnum* (Johnson & Shaw 2015). Of the 91 *Sphagnum* species currently recognized in North America, gametophytes are bisexual in14 and unisexual in 58; the remaining species have unknown sexual conditions (McQueen & Andrus 2007). Bisexual species, as expected, are characterized by high levels of selfing but some outcrossing does occur (Johnson & Shaw 2015). Moreover, mating patterns are related to microhabitat. Among bisexual taxa, hummock species tend to have higher inbreeding coefficients that hollow species, whereas in unisexual species, taxa growing in hummocks high above the water table tend to have lower inbreeding coefficients. Individual female gametophytes generally carry multiple sporophytes (each the result of sexual fusion) and in all species studied to-date, these cohorts of sporophytic offspring are sired by multiple fathers (male gametophytes). Inbreeding depression appears to be rare in *Sphagnum*, even among species with unisexual gametophytes (Johnson & Shaw 2015).

The large number of *Sphagnum* species, their diversity in mating patterns, and their clear patterns of niche differentiation, make the peatmoss clade exceptionally valuable for research in ecological and evolutionary genomics. The *Sphagnum* genome project holds the potential to elucidate underlying genetic factors that determine ecological differences among closely related species, and their evolution. The observation that sporophytes attached to individual maternal gametophytes generally have different fathers opens the possibility for research on genetic imprinting and

other parent-of-origin effects on gene expression.

Peatmosses and global biogeochemistry: linking genes to ecosystem function—Clymo & Haywood (1982) famously commented that there's probably more biomass in *Sphagnum* than in any other genus of plants on earth. *Sphagnum* species dominate many wetlands and produce huge deposits of peat, partially decomposed plant material made up largely of the lower portions of *Sphagnum* plants. Peat accumulates where net production exceeds decomposition over millennia. As a result, northern peatlands are estimated to harbor more than 400 Gt of carbon and are important to global fluxes of CO_2 and CH_4 (Gorham 1991, Yu 2012). Boreal species of *Sphagnum* vary in rates of both production and decomposition and at least some of the physiological and morphological traits that correlate with ecologically important species differences have been identified (Cornelissen et al. 2007, Turetsky et al. 2008, Rydin & Jeglum 2013, Lindo et al. 2013). Peatmoss species also differ in uptake rates for nitrogen, phosphorus and other nutrients that impact photosynthetic rates, as well as interactions with other plants and animals (Hajek & Beckett 2008, Granath et al. 2009).

Sphagnum provides an opportunity to identify plant traits that scale up to impact biogeochemical processes at a global scale (Weston et al. 2015). *Sphagnum*-dominated peatlands are natural laboratories with intra- and interspecific variation at the nucleotide level, and in physiology, morphology, net production, decomposition and carbon accumulation (peat formation). Interspecific variation in a phylogenetic context shows how traits that impact ecosystem function evolved through deep time, and intraspecific variation

can reveal patterns of local adaptation and the action of ongoing natural selection at multiple levels from the genome to phenotypic traits.

Life Cycle – Like all bryophytes, the dominant life-cycle stage for *Sphagnum* is the haploid gametophyte. The gametophyte generation of the life cycle starts with the development of germinating spores that transition from an early filamentous protonematal stage (Fig. 2A) to the thalloid protonematal stage after a few cell divisions (Fig. 2B). This is in contrast to most other mosses, where the protonemata stay filamentous. In *Sphagnum*, the thalloid protonemata forms a leafy gametophore that is the perennial vegetative stage of the *Sphagnum* life cycle. Rhizoids are produced by the thalloid protonemata (Fig. 2B) but are generally absent from mature gametophytes. *Sphagnum* species are typically dioecious, meaning that sperm and eggs are produced on separate gametophytes (and unlike *Physcomitrella*, where individual gametophytes form both sperm and eggs). Mature male *Sphagnum* gametophytes produce antheridia (Fig. 2C) with sperm while female gametophytes produce the egg-containing archegonia (Fig. 2C). As in other mosses, the sperm and eggs are produced mitotically, so they are genetically identical. In nature, the reproductive structures are generally produced once each year, at a predictable season (like flowers). Many boreal species form antheridia and archegonia in the autumn and sporophytes mature in late spring to late summer. Reproductive seasons are species-specific, but with substantial overlap. The sperm require water to reach an archegonium, and swim down the neck of the archegonium to access the egg cell in the archegonial venter. Once the sperm penetrates the egg to form the zygote

and divides to form the embryo, cell specialization occurs to form the foot, which embeds the sporophyte into the maternal gametophyte and the sporangium (capsule; Fig. 2F) in which meiosis occurs. A seta, or capsule stalk, is not formed in *Sphagnum*. In *Physcomitrella* the seta is very short whereas in *Ceratodon*, the other developing moss model, it is well developed (> 1 cm long). The formation of the zygote marks the beginning of the diploid sporophyte stage and at maturity meiosis occurs within the capsule, producing haploid spores. As in other mosses, the sporophyte is green when immature but loses its chlorophyll after meiosis as the spores mature.

The genome project – The Department of Energy (DOE) Joint Genome Institute (JGI) has recently accepted a community supported science project aimed at sequencing two *Sphagnum* species adapted to contrasting microenvironment conditions, and the resequencing of 194 individuals from a *S. fallax* pedigree. The inception of this project provides the first plant-based genome sequencing project aimed specifically at carbon cycling genomics in an ecologically relevant system to enhance our understanding of (1) the genetic variation in natural populations of an undomesticated haploid organism, and (2) genotype-to-phenotype relationships necessary for a trait-based understanding of non-vascular plants in ecosystem function. Furthermore, the resulting genomes from the *Sphagnum* species can be compared to the JGI Plant Flagship model for bryophytes, *Physcomitrella patens* (Rensing et al. 2008), for which there are ample genomic resources available. In comparison to *Physcomitrella* (and *Ceratodon*), *Sphagnum* represents an early diverging lineage of mosses and includes hundreds of

ecologically diverse species, making it an excellent complement that more readily enables research in ecological and evolutionary genomics.

Genomic resources for *Sphagnum* are rapidly progressing. A draft genome sequence (named V0.5, phytozome.jgi.doe.gov/Sfallax_er) for *Sphagnum fallax* (subg. *Cuspidata*) is currently in 49x coverage, assembled from short read Illumina sequencing in 1,228 scaffolds greater than 1kb, with half of the genome on 61 scaffolds that are at least 1.8 Mb in length (L50/N50: 61/1.8Mb). Furthermore, the scaffolds are nearly complete with only 0.4% gap bases between the assembled contigs. There is a robust annotation based on 3.9 billion RNA-seq transcript reads from 16 different growth conditions, resulting in 32,298 genes (26,030 loci; 5,359 alternative splice variants). The main genome is 396 Mb of sequence, the chloroplast genome is in 12 scaffolds at 179 Kb, and the mitochondrial genome is present in 33 scaffolds at 47.6Kb. The *S. magellanicum* genome is currently in progress and appears similarly amenable to genome sequencing. The current draft genome sequence is in 6,255 scaffolds, with half the genome in 464 scaffolds that are at least 294 Kb in length. The main genome, i.e., total scaffold size, is 487.8 Mb. The draft genome sizes for both *Sphagnum* species compare favorably with the genome of *Physcomitrella patens*, whose sequenced genome is 467Mb in length.

A genetic map for *S. fallax* is being produced to generate a chromosome level scale genome assembly for both species. The map is being generated from a 480 member pedigree population that was developed by isolating spores from a single sporophyte, and clonally propagated through single stem descent to ensure individual genotypes. A subset of a 194

individuals from the pedigree population is being used for genome resequencing and trait-based characterization.

Culture and germplasm resources – Establishing *in vitro* growth conditions is essential for the progression of genetic transformation efforts, culturing of organs and tissues, axenic whole plant regeneration, genotype preservation and large-scale biotechnological application (e.g., bioreactors). As noted by Hohe and Reski (2005), the establishment of pure protonematal cultures of *Atrichum undulatum* and *Hypnum velutinum* marked the beginning of a long history of bryophyte *in vitro* culturing. The moss model, *Physcomitrella patens*, benefits from well-established *in vitro* methods that allow for developmental manipulation of the organism, as well as axenic conditions suitable for 'omics' - based assays (e.g., protein profiling Sarnighausen et al. 2004; e.g., metabolites, Erxleben et al. 2012, e.g., RNA-seq, Wu et al. 2014). *Sphagnum* has been grown *in vitro* for quite some time (e.g., Simola 1969), but it is not until relatively recently that a systematic approach was established (Beike et al. 2015). As motivated by the *Sphagnum* genome project and the use of the 'moss bag technique' for biomonitoring as part of the European FP7 "MOSSCLONE" project (www.mossclone.eu), Beike and colleagues (2015) developed and tested *Sphagnum* specific axenic medium and culturing techniques for *S. fimbriatum*, *S. magellanicum*, *S. palustre* and *S. rubellum*. Large-scale bioreactor culturing techniques were also developed and demonstrated for *S. palustre* (Beike et al. 2015). Together with the genome sequences, *Sphagnum* can now benefit from the full suite of molecule 'omics' and systems biology approaches.

In addition to the development of culturing methods, there is a growing number of *Sphagnum* clones being generated and maintained within laboratory conditions. For example, the *S. magellanicum* and *S. fallax* clonal strains used in the DOE JGI sequencing project are available in axenic *in vitro* culture (Fig. 3A; Weston Lab, Oak Ridge National Lab. USA). Furthermore, a 480 - member pedigree derived from a single sporophyte is available in tissue culture plates and are amenable to high-throughput phenotyping for growth morphology and chlorophyll fluorescence assays (Fig. 3B, C).

Challenges and opportunities with population and genomic resources – While the development of *Sphagnum in vitro* culturing techniques, germplasm resources and draft genome assemblies marks the beginning of its use for ecological and evolutionary genomics, there are still substantial obstacles to be overcome before the utility of these resources are fully realized. For example, we are not able to manipulate sexual reproduction within laboratory conditions or perform controlled crosses that would greatly strengthen current approaches for linking traits to underlying quantitative trait loci (QTL). Furthermore, we have not yet perfected the transfer of *Sphagnum* from *in vitro* axenic cultures to field conditions where so many of our ecological and evolutionary questions reside. Also lacking is a stable nuclear transformation protocol that is present in most plant model systems. In *P. patens*, gene targeting through homologous recombination has been demonstrated (Kamisugi et at. 2006). In this approach, gene replacement occurs at a desired chromosomal location (locus), thereby alleviating the negative effects of undirected and multiple gene copies often

observed through biolistics (gene gun technology), or delivery from *Agrobacterium*.

Although challenges in developing this emerging model system exist, there are considerable benefits in using *Sphagnum* to decipher genotype-to-phenotype relationships. For example, the ease by which large numbers of individuals can be grown, stored, assayed and phenotypically characterized is a major benefit of using bryophytes, and *Sphagnum* specifically. In most crop plants and bioenergy feedstocks, phenotype characterization including the structural, physiological and performance-related traits of genotypes in a given environment is a considerable challenge (Benfey & Mitchell-Olds 2008; Dhondt et al. 2013). This may not be a major constraint for *Sphagnum* as it can be readily grown in laboratory culture with multiple media and environmental conditions. The 480-member *S. fallax* population discussed above is stored in a series of 12-well tissue culture plates that occupy 1.2 m^2 of illuminated shelf space (Fig. 3B). This in essence, represents the 'common garden' that requires considerably less resources in space and cultural practices relative to crop plants and bioenergy feedstocks. Furthermore, the plate-based system is amenable to phenotyping assays that include growth analyses and characterization of photosynthetically related traits such as chlorophyll-*a* fluorescence and growth responses to substrate pH (Fig. 4). The ability to store entire populations of *Sphagnum* in axenic *in vitro* culture coupled with high-throughput imaging and image processing for trait characterization is a tremendous compliment to the population – based sequencing efforts that will facilitate our understanding of genotype-to-phenotype relationships.

An evolutionary genomics example: epigenetics and Sphagnum reproduction – Several features of moss (including *Sphagnum*) mating systems make them especially valuable for studies of reproductive biology (Johnson & Shaw 2015). First, mosses (like liverworts and hornworts) are the only land plants in which the fitness of gametophytes is a quantitative trait. This is true because moss gametophytes can parent more than one sporophytic offspring, and a single gametophyte produces many genetically identical eggs because the gametes are produced mitotically. As a result of gametophyte clonal reproduction, genetically identical eggs and sperm can participate in many independent fertilization events. In all other plants, including the spore-producing ferns, gametophytes parent either one sporophyte or none; i.e., fitness is binary. The second feature is that multiple paternity for sporophytes attached to an individual female gametophyte appears to be the rule in *Sphagnum*, and probably in other mosses. Multiple paternity sets up the conditions in which parent-offspring conflicts can be expected, and indeed the mosses represent an ideal group for studying this phenomenon (Haig & Wilczek 2006). Parent-of-origin effects on gene expression in offspring are well-known in angiosperms (and mammals), but it is easier to study in mosses because in angiosperms, seeds include complex combinations of tissues of maternal and paternal origin, as well as triploid endosperm with unequal contributions from the two parental genomes.

We describe here a preliminary analysis of (gametophytic) parental effects on (sporophytic) gene expression in *Sphagnum palustre*. This species is an allopolyploid (diploid gametophytes, tetraploid sporophytes), which

limited our ability to unambiguously identify paternal alleles in sporophytes, but the analysis shows proof of concept for future studies.

Two female gametophytes each bearing multiple sporophytes were sampled from a natural population (West Virginia). RNA was extracted from the two maternal gametophytes, and from a total of seven attached sporophytic offspring (three from one gametophyte, four from the other) using RNAzol®RT (Molecular Research Center, inc.). RNA-seq (Wang et al. 2009) paired-end libraries were constructed using the Kapa stranded mRNA library preparation kit (Kapa Biosystems). Libraries were indexed, pooled and sequenced on two lanes of an Illumina HiSeq 2500 sequencer flow cell (Illumina, San Diego, California, USA). Sequencing generated 100pb paired-end sequences.

The sequence data were assembled for the two maternal gametophytes using Trinity_r20131110 (Grabherr et al. 2011, Haas et al. 2013). In order to remove potential contaminant transcripts (i.e., from contaminating organisms) from the assemblies, the transcriptome assemblies were blasted against the Uniprot database (The UniProt Consortium 2014) using BLASTX (Altschul et al. 1997). BLASTX outputs were filtered to discard transcripts with top hits from non-land plants. For the two maternal gametophytes, we obtained 78040 and 8735 hits to the *Physcomitrella* proteome and another 2992 and 4510 hits to land plants other than *Physcomitrella*. The expression level of each RNA unit was measured by the number of sequenced fragments that mapped to the transcript, which is expected to correlate directly with its abundance level. One of the gametophytic transcriptomes was used as a reference to which all the

sequences from the various sporophytic RNA samples were mapped using Bowtie (Langmead et al. 2009). The reads from the gametophytes were also mapped to this reference transcriptome. RSEM (Li & Dewey 2011) was used to produce the gene count matrix for all differential expression analyses. DESeq was used on to analyze the count data and test for differential expression between gametophytes and sporophytes and between sporophytes form one maternal gametophyte versus the other (parental effect).

Gene expression profiles from the two maternal gametophytes and seven sporophytes indicate expression differences between sporophytes and gametophytes; a total of 2420 genes exhibited significant expression differences between the two life cycle generations (Fig. 5A). This is also indicated by clustering the nine samples (two gametophyte, seven sporophyte) by similarities in overall expression profiles (Fig. 5B). Moreover, differential expression of genes occurred among sporophytes borne on the two different maternal gametophytes, indicating epigenetic effects associated with those gametophytes on the sporophytes they parent (Fig. 5C). This is also indicated by multigenic expression profiles, although two of the sporophytes from gametophyte M1 group with sporophytes from M2 (Fig. 5B).

The results of this preliminary experiment reveal substantial expression differences between sporophytes and gametophytes in *S. palustre*. Parent-of-origin effects on gene expression are strongly suggested although we cannot determine whether these effects reflect maternal genetic effects, paternal genetic effects, or maternal environmental effects. The

experiment nevertheless demonstrates the utility of *Sphagnum* for investigating epigenetic effects on gene expression and provides a valuable system for investigating mating systems at the genomic level.

Synthesis: the potential for a genus level sequencing project – Whereas the *Ceratodon* and *Physcomitrella* models have been developed primarily for intensive evo-devo investigations, *Sphagnum's* particular value is in ecological and evolutionary genomics. The *Sphagnum* system facilitates adding new dimensions to genomic research through comparative studies of ecologically diverse species in a well-studied clade. *Physcomitrella* is not common or abundant in nature. The genus Physcomitrella was traditionally thought to include three or four species (http://www.tropicos.org/NameSearch.aspx?name=Physcomitrella&commonname=) but phylogenetic analyses (Medna et al. 2015) indicate that there are only two species, and they do not form a monophyletic group. *Ceratodon* is also a genus of only about four species worldwide, though *C. purpureus* is common around the Northern Hemisphere. The fact that peatmosses have such profound impacts on global biogeochemistry and therefore climate provides exceptional added value for functional ecology and ecological genomics.

We envision a research program in which *Sphagnum* species representing the phylogenetic and ecological diversity of peatmosses are characterized at the genomic and phenotypic levels to determine how genome structure and function translate into plant traits, and through those traits, to niche differentiation and patterns of carbon cycling. Species to be

included should include taxa that inhabit nutrient poor bogs, richer fens, those that form hummocks high above the water table and accumulate a lot of peat, and fully aquatic species that differ in relevant traits and decompose faster. Boreal, tropical, and Southern Hemisphere species open a more macro dimension to comparative analyses of ecology, and inclusion of species representing the major clades within *Sphagnum* can reveal evolutionary changes over deep time scales. Based on molecular data (Shaw et al. 2010b) the Sphagnopsida likely diverged from other mosses 100-200 mya, but *Sphagnum* s. str. appears to have diversified rapidly during the Neogene and this agrees closely with estimates of when *Sphagnum*-dominated wetlands appear in the fossil record (Greb et al. 2006). Inclusion of species from the related genera *Ambuchanania, Eosphagnum*, and *Flatbergium* in future genomic work will (potentially) enable a complete picture of how peatmosses became wetland dominants as global climate cooled during the Tertiary and boreal peatlands came into existence.

Acknowledgements

The time and resources to produce this chapter were supported by U.S. Department of Energy, Office of Science, Biological and Environmental Research, including the SPRUCE project (http://mnspruce.ornl.gov/) and the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725. The work conducted by the U.S. Department of Energy Joint Genome Institute was supported by the Office of Science of the U.S.

Department of Energy under Contract No. DE-AC02-05CH11231.

References

Andrus, R. E. (2006). Six new species of *Sphagnum* (Bryophyta: Sphagnaceae) from North America. *Sida,* 22, 959-972.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389-3402.

Beike, A. K., Spagnuolo, V., Lüth, V., Steinhart, F., Ramos-Gómez, J., Krebs, M., Adamo, P. (2015). Clonal in vitro propagation of peat mosses (*Sphagnum* L.) as novel green resources for basic and applied research, *Plant Cell Tissue Organ Cult*.,120,1037-1049.

Benfey, P..N., Mitchell-Olds, T. (2008). From genotype to phenotype: systems biology meets natural variation. *Science*, 320, 495-497.

van Breemen, N. (1995). How *Sphagnum* bogs down other plants. *Trends in Ecology & Evolution,* 10, 270–275.

Bower, F. O. (1890). On antithetic as distinct from homologous alternation of generations in plants. *Annales of Botany,* 4, 347-370.

Chiang, Y., Graham, S.W. (2011). Inferring the higher-order phylogeny of mosses (Bryophyta) using a large, multigene plastid dataset. *American Journal of Botany*, 95, 839-849.

Clymo R. S., Hayward P. M. (1982). *The Ecology of Sphagnum*. UK and New York, NY, USA: Chapman and Hall Ltd., London.

Cornelissen J.H.C., Lang S.I., Soudzilovskaia N.A., During H.J. (2007). Comparative cryptogam ecology: a review of bryophyte and lichen traits that drive biogeochemistry. *Annals of Botany,* 99, 987–1001.

Dhondt, S., Wuyts, N., Inzé, D. (2013) Cell to whole-plant phenotyping: the best is yet to come. *Trends in Plant Science,* 18, 428-439.

Erxleben, A., Gessler, A., Vervliet-Scheebaum, M., Reski, R. (2012). Metabolite profiling of the moss *Physcomitrella patens* reveals evolutionary conservation of osmoprotective substances. *Plant Cell Reports,* 31, 427-436.

Floyd, S.K., Bowman, J.L. (2007). The ancestral developmental tool kit of land plants. *International Journal of Plant Sciences, 168*, 1-35.

Gorham E. (1991). Northern peatlands: role in the carbon cycle and probable responses to climatic warming. *Ecological Applications,* 1, 182–195.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J.Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. et al. 2001. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology*, 29, 644-652.

Graham, L. E., Kodner, R. G., Fisher, M. M., Graham, J. M., Wilcox, L. W., Hackney, J. M., Obst, J., Bilkey, P. C., Hanson, D. T., Cook, M. E. (2003). Early land plant adaptation to terrestrial stress: a focus on phenolics (pp. 155–171). In: A. Hemsley, I. Pecie, (Eds.), *The Evolution of Plant Physiology*. London: Academic Press.

Granath, G., Wiedermann, M.M., Strengbom, J. (2009) Physiological responses to nitrogen and sulphur addition and raised temperature in *Sphagnum balticum*. *Oecologia,* 161, 481–490.

Greb, S. F., DiMichele, W. A., Gastaldo, R. A. (2006). Evolution and importance of wetlands in earth history. In Greb, S. F., DiMichele, W. A. (Eds.), Wetlands through time: Geological Society of America Special Paper 399, p. 1-40,

Haeckel, E., (1868). *The History of Creation*. New York: Appleton and company.

Haig, D., Wilczek, A. (2006). Sexual Conflict and the alternation of haploid and diploid generations. *Philosophical Transactions: Biological Sciences,* 361, 335–343.

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith Jr, R. K.,Jr., Hannick, L. I. et al. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.,* 31, 5654-5666].

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols,* 8, 1494-1512.

Hajek, T., Beckett, R. P. (2008) Effect of water content components on desiccation and recovery in Sphagnum mosses. *Annals of Botany* 101, 165–

173.

Hohe, A., Reski, R., (2005) From axenic spore germination to molecular farming - one century of bryophyte in vitro culture. *Plant Cell Rep.,* 23, 513–521.

Johnson, M. J., Shaw, A.J., (2015). Genetic diversity, sexual condition, and microhabitat preference determine mating patterns in *Sphagnum* (Sphagnaceae) peat-mosses. *Biological Journal of the Linnean Society,* 11, 96–113.

Johnson, M. J., Granath, G., Teemu, T., Pouliot, R., Stenøien, H. K., Rochefort, L., Rydin, H., Shaw, A. J. (2015). Evolution of niche preference in *Sphagnum* peat mosses. *Evolution,* 69, 90-103.

Kamisugi, Y., Schlink, K., Rensing, S. A., Schween, G., von Stackelberg, M., Cuming, A. C., et al. (2006). The mechanism of gene targeting in *Physcomitrella patens*: homologous recombination, concatenation and multiple integration. *Nucleic Acids Research*, *34*(21), 6205–6214.

Karlin, E. F., Boles, S. B., Ricca, M., E. Temsch, M., Grelihuber, J., Shaw, A.J. (2009). Three genome mosses: complex double allopolyploid origins for triploid gametophytes in *Sphagnum*. *Molecular Ecology,* 18, 1439–1454.

Kendrick, P., Crane, P.R. (1997) The origin and early evolution of plants on land. *Nature,* 389, 33–39.

Knack, J. J., Wilcox, L. W., Delaux, P.-M., Ané, J.-M., Piotrowski, M. J., Cook, M. E., Graham, J. M., Graham, L.E. (2015). *International Journal of Plant Science*, 165, 405-420.

Kyrkjeeide, M.O., Hassel, K., Stenøien, H.K., Prestø, T., Boström, E., Shaw, A.J. , Flatberg, K.I. (in Press). The dark morph of *Sphagnum fuscum* (Schimp.) H.Klinggr. in Europe is conspecific with the North American *S. beothuk*. Journal of Bryology

Langmead, B., Trapnell, C., Pop, M., Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology,* 10, R25.

Li, B., Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics,* 12, 323.

Lindo, Z., Nilsson, M. C. & Gundale, M. J. (2013). Bryophyte-cyanobacteria associations as regulators of the northern latitude carbon balance in response to global change. *Global Change Biology,* 19, 2022–2035.

McQueen, C.B., Andrus, R.E. (2007). Sphagnaceae, pp. 45–101. In Flora of North America North of Mexico (ed. C. FNAE). New York, NY, USA and Oxford, UK.

Qiu, Y. L., Li, L., Wang, B., Chen, Z. Knoop, V., Groth-Malonek, M. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences USA,* 103, 15511–15516.

Redecker, D., Kodner, R., Graham, L.E. (2000). Glomalian fungi from the Ordovician. *Science,* 289, 1920-1921.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., et al. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science,* 319, 64–69

Ricca, M., Shaw, A.J. (2010). Allopolyploidy and homoploid hybridization in the *Sphagnum subsecundum* complex (Sphagnaceae: Bryophyta). *Biological Journal of the Linnaean Society,* 99, 135–151.

Rubinstein, C. V., Gerrienne, P., de la Puente, G. S., Astini, R. A., Steemans, P. (2010). Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana). *New Phytologist,* 188, 365–369.

Rydin H., Jeglum J. (2006). *The Biology of Peatlands*. New York, NY, USA: Oxford University Press.

Sarnighausen, E., Wurtz, V., Heintz, D., Van Dorsselaer, A., Reski, R. (2004). Mapping of the *Physcomitrella patens* proteome. *Phytochemistry*, 65, 589-1607

Smit, A.F.A., Hubley, R., Green, P. *RepeatMasker Open-3.0.* 1996-2011 <http://www.repeatmasker.org >.

Shaw, A. J., Cox, C. J. Buck, W. R., Devos, N., Buchanan, A. M., Cave, L., Seppelt, R., Shaw, B., Larraín, J., Andrus, R. E., Greilhuber, J., Temsch, E.M. (2010a). Newly resolved relationships in an early land plant lineage: Bryophyta class Sphagnopsida (peat mosses). *American Journal of Botany*, 97,1511–1531.

Shaw, A. J., Devos, N., Cox, C. J., Boles, S. B., Shaw, B., Buchanan, A. M., Cave, L., Seppelt, R. (2010b). Peatmoss (*Sphagnum*) diversification associated with Miocene Northern Hemisphere climatic cooling? *Molecular Phylogenetics and Evolution*, 55, 1139–1145.

Simola, L.K. (1969) The effect of various mono- and disaccharides on the growth of *Sphagnum nemoreum* thalli in sterile cultures. *Physiologia Plantarum*, 22, 1079–1084.

Steemans, P., Hérissé, A. L., Melvin, J., Miller, M. A., Paris, F. Verniers, J. et al. (2009). Origin and radiation of the earliest vascular land plants. *Science,* 324, 353.

Turetsky M. R., Bond-Lamberty, B., Euskirchen, E., Talbot, J., Frolking, S., McGuire, A. D., Tuittila, E. S. (2012). The resilience and functional role of moss in boreal and arctic ecosystems. *New Phytologist* 196: 49–67.

Turetsky, M. R., Crow, S. E., Evans, R. J., Vitt, D. H., Wieder, R. K. (2008) Tradeoffs
in resource allocation among moss species control decomposition in boreal peatlands. *Journal of Ecology,* 96, 1297–1305.

Viaene, T., Landberg, K., Thelander, M., Medvecka, E., Pederson, E., Feraru, Cooper, E.D., Karimi, M., Delwiche, C.F., Ljung, K., Geisler, M., Sundberg, E., Friml, J.. (2014). Directional auxin transport mechanisms in early diverging land plants. *Current Biology*, 24, 2786-2791.

Vitt, D. H., Slack, N. G. (1975). An analysis of the vegetation of *Sphagnum*-dominated kettle-hole bogs in relation to environmental gradients. *Canadian Journal of Botany*, 53, 332-359.

Wang, Z., Gerstein, M., Snyder M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics,* 10, 57-63.

Waters, E. R. (2003). Molecular adaptation and the origin of land plants. *Molecular Phylogenetics and Evolution,* 29, 456–463.

Wellman, C. H., Osterloff, P. L., Mohiuddin, U. (2003). Fragments of the earliest land plants. *Nature,* 425, 282–285.

Weston, D. J., Timm, C. M., Walker, A. P., Gu, L., Muchero, W., Schmutz, J., et al. (2015) *Sphagnum* physiology in the context of changing climate: emergent influences of genomics, modelling and host–microbiome

interactions on understanding ecosystem function. *Plant, cell & environment,* 38, 1737-51.

Wickett, N. J., Mirarab, S., Nguyen, N., Carpenter, E., Matasci, N., Ayyampalayam, S., et al. (2014). Phylotranscriptomics analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences USA,* 11, 4859-4868.

Wu, H. P., Su, Y. S., Chen, H. C., Chen, Y. R., Wu, C. C., Lin, W. D., Tu, S. L. (2014). Genome-wide analysis of light-regulated alternative splicing mediated by photoreceptors in *Physcomitrella patens. Genome Biology,* 15, R10.

Yu, Z. C. (2012). Northern peatland carbon stocks and dynamics: a review. *Biogeosciences*, 9, 4071–4085.

Figure Legends:

Fig. 1. Niche differentiation among *Sphagnum* species in terms of distance from the mire margin. A similar sequence can be observed within the mire relative to the hummock-hummock gradient. *Sphagnum teres* occupies relatively mineral-rich microhabitats near the mire margin. Modified from Vitt & Slack (1975). In the original figure, *S. capillifolium* was labeled *S. capillacium*, a later name sometimes used for this species; *S. angustifolium* was labeled *S. recurvum*, a name sometimes applied to a group of closely related species including *S. angustifolium*, to which the plants included in that study likely belonged.

Fig. 2. *Sphagnum* life cycle. A. Young protonema with a single rhizoid. B. Mature thalloid protonemata with rhizoids. C. Male gametophore. D. Female gametophore. E. Female gametophyte bearing multiple sporophytes. F. Mature sporophyte (round, brown capsule, bearing meiotically-derived spores) borne on a short pseudopodium (mostly surrounded by perichaetial leaves.

Fig. 3. Representative cultivation techniques for *Sphagnum*. Gametophytes can be cultivated on solid medium in (A) magenta vessels, (B) in petri dish, (C) populations in a series of multi-well tissue culture plates, or (D) liquid Knop medium in a photobioreactor (Photo credit for panel D: Anna Beike, Reski Lab, University of Freiburg, Germany).

Fig. 4. Photosynthesis and growth phenotypic distribution of Sphagnum cultivated for 3 weeks at pH 4.5 (red) and 8.5 (blue). (A) maximum quantum yield of PSII, (B) percent growth (plant area after three weeks - initial plant area)/initial plant area * 100).

Fig. 5. Gene expression in gametophytes and sporophytes of *S. palustre*. A. Volcano plot showing genes (grey) that are differentially expressed (at $P<0.0001$) in sporophytes vs. gametophytes. B. Heatmap showing cluster analysis of sporophyte and gametophyte samples. M1 and M2 = maternal gametophytes; S1-S4 = sporophytes attached to M1 or M2. C. Volcano plot showing genes (grey) that are differentially expressed (at $P<0.0001$) in sporophytes produced by maternal gametophyte M1 vs. M2.